AD-A105 079 FORD AEROSPACE AND COMMUNICATIONS CORP SUNNYVALE CA E--ETC F/G 5/9
AN INFORMATION PROCESSING MODEL FOR PREDICTING ERROR DETECTION --ETC(U)
JUL 81 R A GOLDBECK, F M FERRANTE F49620-79-C-0147
UNCLASSIFIED ESD-SCFO-TR6210 AFOSR-TR-81-0681 NL

1 OF 1
ADA
105079

END
DATE
FILMED
10-81
DTIC

AFOSR-TR- 81 -0681

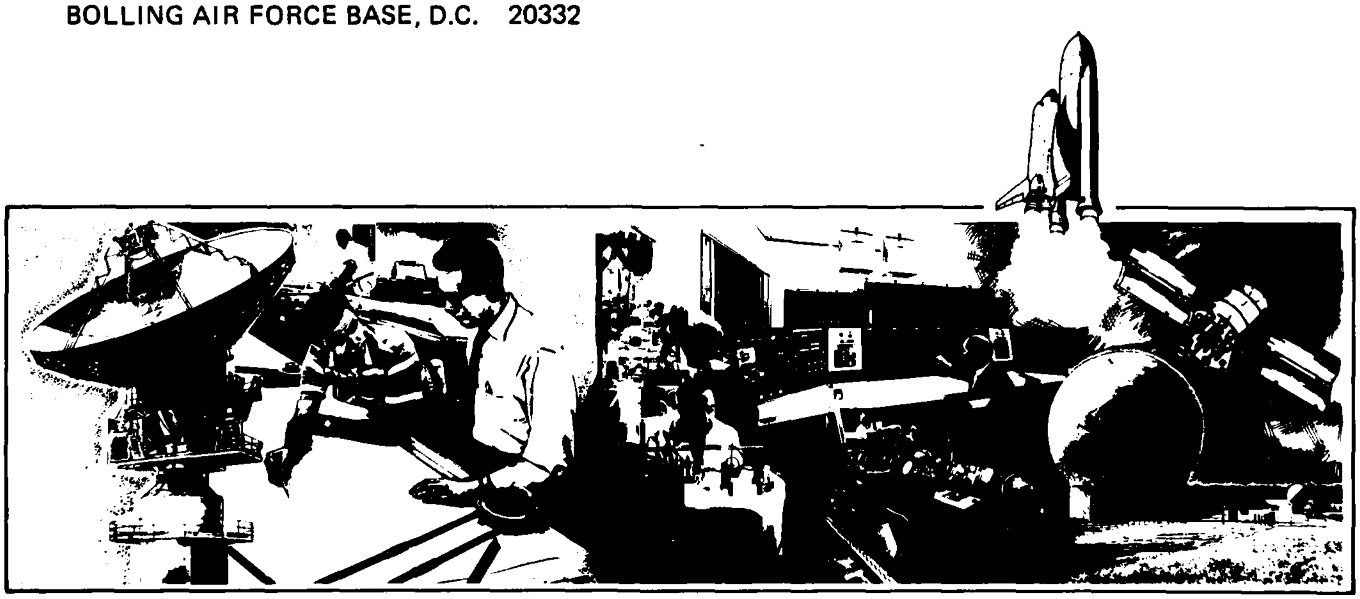
ESD-SCFO-TR6210
July 1981

AD A105079

AN INFORMATION PROCESSING MODEL FOR PREDICTING ERROR DETECTION AND CORRECTION: AN ANALYSIS OF OPERATOR PERFORMANCE

Prepared for:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
AIR FORCE SYSTEMS COMMAND, USAF
BOLLING AIR FORCE BASE, D.C. 20332



Ford Aerospace &
Communications Corporation
Engineering Services Division
1260 Crossman Avenue
Sunnyvale, California 94086

81 10 2 110

Ford Aerospace & Communications Corporation

ESD-SCFO-TR6210

AN INFORMATION PROCESSING MODEL FOR PREDICTING ERROR DETECTION AND CORRECTION: AN ANALYSIS OF OPERATOR PERFORMANCE

Robert A. Goldbeck
Felice M. Ferrante

Final Report for Period: June 1979 - May 1981
21 July 1981

Final Scientific Report
Prepared under contract with the
Air Force Office of Scientific Research
Air Force Systems Command, USAF
Bolling Air Force Base, D. C. 20332

FORD AEROSPACE & COMMUNICATIONS CORPORATION
ENGINEERING SERVICES DIVISION
SUNNYVALE, CALIFORNIA

Acknowledgements

The research described in this report was conducted by the Ford Aerospace & Communications Corporation, Engineering Services Division, Sunnyvale, California, under Contract Number F49620-79-C-0147, with the Air Force Office of Scientific Research, Air Force Systems Command, United States Air Force. Alfred Fregly was the Program Manager for the Life Sciences Directorate, AFOSR and Donald Topmiller was the Technical Monitor from Aerospace Medical Research Laboratories.

Many people have contributed their time and efforts on this research endeavor. The authors are pleased to extend acknowledgement to the following contributors. Appreciation is extended to Martin Deggeller and Arthur Walberg for their helpful advice on technical and administrative matters. Appreciation is extended to Jeffrey Charlet who assisted in the early phases of this project, as well as lending insight and suggestions throughout the project duration. Special thanks go to Keith Walberg, who performed the display programming for this project and Karen Kimura for her statistical package programs. Appreciation is extended to Rudy Gath for the hardware development. Finally the authors wish to thank Kim Zee for her patience and typing skills in the preparation of this report.

TABLE OF CONTENTS

List of Figures

List of Tables

SECTION 1

INTRODUCTION

RESEARCH OBJECTIVES

The overall objective of this research was to develop the capability to predict operator performance at a system console. To be useful in industry, this prediction should be available prior to production of a prototype or simulator. The value of such predictions is the opportunity for an analyst to select a design that meets system performance requirements while optimizing human performance -- and to be able to make this selection before hardware and software design and development activities commence. For this study, three models were examined to determine how each could contribute to the formulation of an integrated model for predicting human performance.

An additional objective was the capability of predicting human error detection and correction probabilities. Human errors that are not detected and corrected by the operator become system errors which may impact system reliability. Human errors that are detected and corrected remain as human errors, without system reliability impact. The act of detection and correction of errors, however, can be expected to increase system performance time. For a task that has stringent time performance requirements, error detection and correction may not be possible and thus, system errors may result.

BACKGROUND

A computer program entitled Human Engineering Computer Aided Design (HECAD) was developed by AMRL. This program was designed to predict operator task times and error rates at operator workstations. The program uses files of the AIR Data Store (Munger, Smith, & Payne, 1962) to obtain times and reliabilities of control activations and display readings; and methods-time-measurement to obtain times for hand and eye transfers.

Ford Aerospace, in cooperation with AMRL, conducted a laboratory test of HECAD to validate and update HECAD predictions (Goldbeck & Charlet, 1975). Videotaped hand and eye movement data were used to modify the elements of task time. Also, a distinction was made between link and node errors; where going to the wrong control or display is a link error, and mismanipulating a control or misreading a display is a node error. One of the limitations of HECAD is that it does not predict error detection and correction, so that its output is in human error per se, and not the effect of human error on system performance.

Independent of work on HECAD, Warren Teichner, under contract with ONR and AFOSR at New Mexico State University (Teichner, 1974; Teichner & Williams, 1979), had developed an information transfer theory of performance. His theory proposed at least four stages of processing: (1) Stimulus acquisition, (2) S-S translations (the processes by which the identified stimulus is translated to a new code), (3) S-R translations (translations of a stimulus code to a response code), and (4) Response executions. The S-S and S-R translations are comparable to a HECAD link. Since HECAD link types were used for HECAD prediction of link errors, Teichner's translations are the corresponding predictors for the Teichner theory. Like the HECAD model, the Teichner model does not address error detection and correction.

APPROACH

The relationship between the HECAD and Teichner models is important to arrive at an integrated prediction model. Points of correspondence must be established so that this integrated prediction model can be developed for use. A third component of the integrated prediction model is the Information Metric which is an expression of central processing uncertainty. In order to begin developing the prediction model, two components, or tasks, must be constructed. The first required task is a Data Acquisition task. The major purpose of this task is to identify and acquire the types of data necessary to establish the prediction model. The second task is central to the research strategy for evaluating the model comprised of the three dimensions listed above. This second task is called a Validation task. The purpose of the Validation task is to study how the prediction model, based on data from the Data Acquisition task, can be applied and tested with regard to human performance predictions. After these data are collected, it is then possible to make predictions for the Validation task. Data is then collected from the Validation task, so that the performance predictions and the prediction model can be evaluated.

SECTION 2

METHOD

SUBJECTS

A total of 20 men and women between the ages of 19-33 participated as subjects in the present study. All subjects met the following criteria: no previous experience with control panel operations, no physical defects which would impair their performance on the experimental equipment, 20/20 normal vision (corrected or uncorrected), English as a primary conversational language, and the successful completion of at least one semester in an accredited junior college, college, or university. All subjects were supplied by a temporary employment agency and were paid for their participation.

APPARATUS

In order to present the experimental display and task scenarios to the subjects the following system configuration was used (See Figure 1): (1) a microcomputer with standard peripherals, (2) a display generator with interfaces to the computer system, (3) a cathode-ray tube (CRT) display monitor, and (4) a touch entry device. A brief description of these hardware items follows.

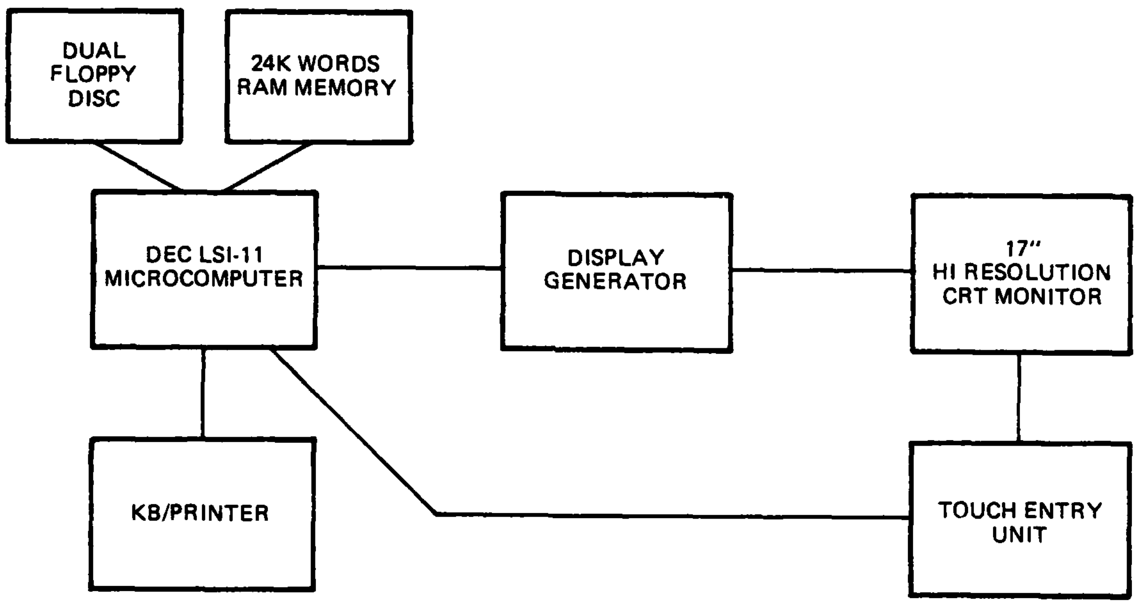


Figure 1. System Configuration

Computer. A 16-bit DEC LSI-11 digital microcomputer and line printer were used to generate both display formats as well as to record and store subject time and error responses.

Display generator. The display generator, developed by FACC, is a high resolution, raster scan display unit capable of generating characters in a 7 x 9 pixel format (9 x 9 font). In addition, the display generator was capable of presenting lines to define a boundary around each switch area used for responses on the touch-entry CRT. When this switch area, or matrix box, was selected by the user, the inside of the box went to reverse video to indicate that a specific area had been selected. Only one switch area selection could be activated or be in reverse video at any given time.

Display monitor. A 17" Conrac CRT, (Model No. RQB 17/N, P4 phosphor), was used to display all stimuli. This CRT is a high resolution (721 X 826 lines by pixels) black and white display with a character size of 7 x 9 pixels. The CRT has a character brightness of 50 ft. L and a contrast ratio of 4:1.

Touch entry device. The touch entry device, manufactured by Carroll Manufacturing, was mounted around the periphery in front of the CRT surface. A total of 88 LED emitters and phototransistor detectors were mounted 1/4" apart to produce a beam matrix of 48 (horizontal) x 40 (vertical). Thus, an 11 3/4" x 9 3/4" area of the CRT was capable of acting as a touch entry switch matrix, with the intersection of the horizontal and vertical beams constituting a touch point. Interruption of an infra-red horizontal or vertical LED beam by the operator's finger resulted in sending a specific CRT coordinate to a temporary storage register located in the interface. This coordinate was then checked by the software to insure that it was a valid point, i.e., located within a particular box of the display. Once validated, the coordinate was processed and transmitted to the main computer for storage on a floppy disk. Operator feedback of switch activation was provided by a 15 msec tone and the selected box area going to reverse video when the CRT screen was touched.

Visual display. All subjects were seated directly in front of the CRT so that the center of the display matrix was approximately 20 deg below the line of sight and all matrix boxes could be easily touched by the operator. The subject's console was positioned in such a way that the experimenter could observe all actions through a one-way mirror.

The visual display was composed of four elements; namely, a matrix area, a variable status field, a time out error box, and a switch error box. The matrix portion of the display area consisted of 36 boxes arranged in a 6 x 6 (row x column) format (See Figure 2). Each box or matrix area was covered by three horizontal and three vertical active beams. The intersection of these beams constituted a valid touch point, thus sending a coordinate to the computer when the inner portion of the box was touched. Each matrix area box was separated by two non-active beams. If the subject

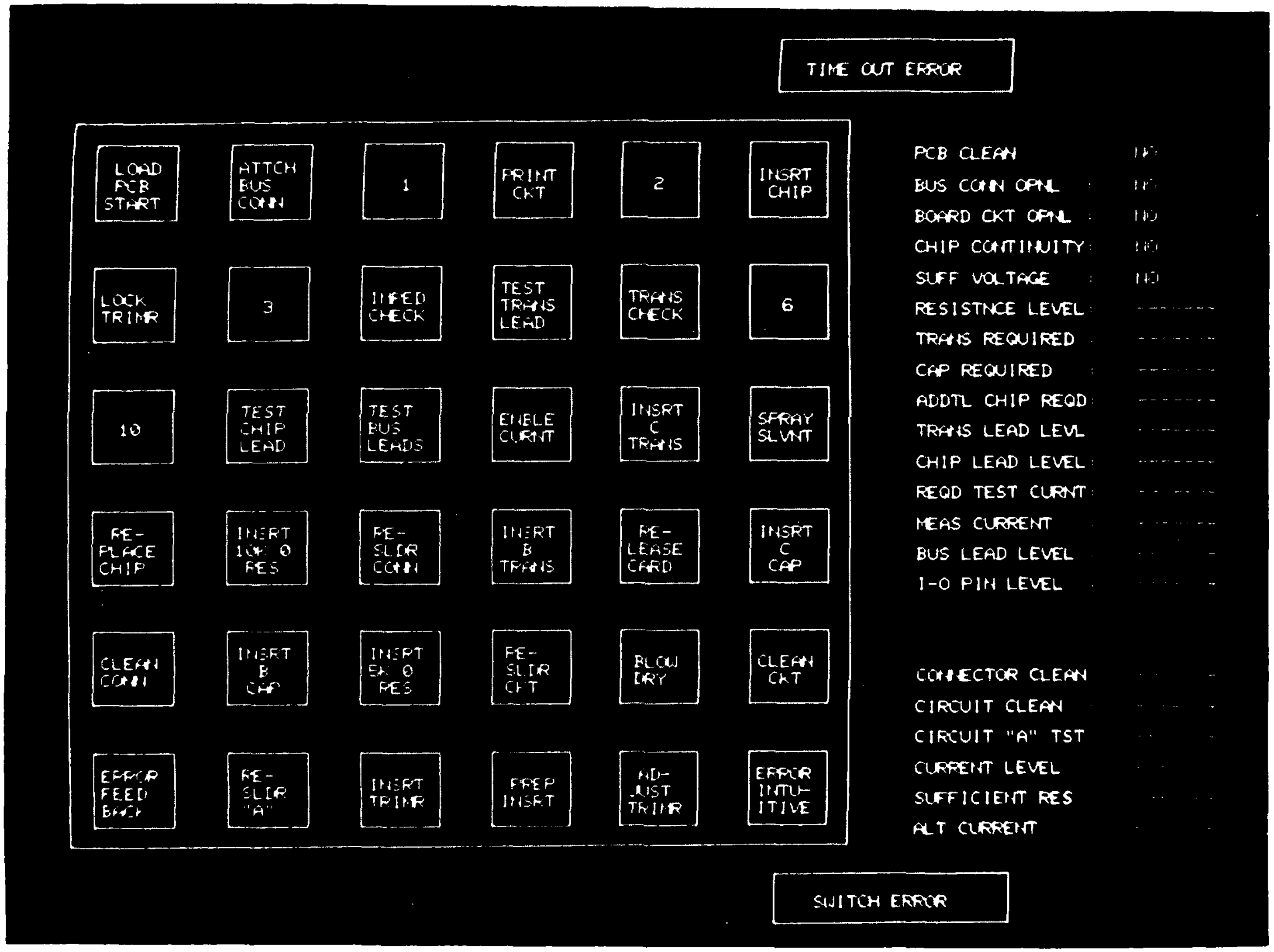


FIGURE 2. MATRIX DISPLAY.

touched any part of the display outside of the matrix boxes, that coordinate was considered invalid and was not sent to the computer.

Thirty four of the matrix boxes corresponded to the control actions necessary to complete both tasks. The layout for the matrix area was designed in such a way that control responses that were correct for the given state of a display item were not located adjacent to the correct control response for a different state of the same display item. This design concept was followed so that an accidental activation of a matrix box could be identified as a node error rather than the more cognitive link error.

Located to the right of the matrix area was a 21 item variable display status field. At the start of each new task segment, each item in the status field had a dashed line opposite it. The dashed lines were eventually replaced by the words yes, no-go, or marginal when a corresponding matrix area was activated, thus giving the operator a current and up-to-date status of the system. Unlike the matrix area boxes, the status field display items were arranged sequentially from top to bottom.

The third element of the visual display was a time out error box. This was located above the matrix area and status field. The time out error light acted as a prompt to get subjects to make a matrix selection when too much time was being taken between two consecutive actions. Subjects were given a total of 20 sec between correct matrix selections. After 10 sec had elapsed between actions the time out error box would briefly go to reverse video. This was basically to give the subject a cue as to how much time had elasped. If too much time was still taken, and a correct matrix box was not selected, the time out error box would go to reverse video and the matrix area would lock up, meaning that no matrix selections could be activated or sent to the computer. When this occurred, the experimenter would give the subject the next correct action and restart the program.

The final element of the visual display was the switch error box. This was located below the matrix area and status field. The switch error box served as a feedback cue to inform the subject that an error had been made. This part of the display would go to reverse video only when two consecutive incorrect matrix box areas were selected. The switch error light remained in reverse video until a correct matrix selection was made.

Error keys. Within the matrix area two boxes were non-task related; namely, the ERROR FEEDBACK key and the ERROR INTUITIVE key. These boxes were located in the lower left and right portion of the matrix display area. Subjects were instructed that they were to use one of the error keys whenever an incorrect matrix switch was activated. The instructions informed the subject to use the ERROR FEEDBACK key when one of the following conditions occurred: the switch error light went to reverse video, the equipment locked up and the experimenter gave the next correct response, or the subject made an incorrect matrix activation and did not realize it until he obtained additional information from the status field. Similarly, subjects were instructed to use the ERROR INTUITIVE key for the following conditions: the wrong matrix switch was accidently activated, or the subject immediately realized that an error was made.

In order to ensure that all subjects fully understood the proper use of each error key, a list of possible error conditions was given to each subject. Subjects were asked to categorize, by means of a true-false quiz, each error condition by the error key they would use for that given situation. Immediately after completing this categorization, the experimenter went over each error condition again with the subject to ensure that the distinction between the two keys and when to use them was clear.

PROCEDURE

The training procedure for both tasks was twofold. The first part of the training consisted of the actual learning and memorization of the task sequences. This was accomplished through the use of detailed tapes outlining the correct actions and manipulations necessary to perform the specific task. The second part of the training procedure provided the subjects with the opportunity for practice and memorization of the task while becoming familiar with the actual experimental equipment. Further elaboration on the training procedure follows.

Training tapes. The training tapes for each task were divided into three segments, namely, a go condition, a no-go condition, and a marginal condition. Two tapes accompanied each condition. The first tape presented each subject with a detailed step by step account of the task to be performed, as well as the rationale for taking each action. The second tape listed only the control and display items the subject needed to activate or check for each condition; all rationale was omitted. In the construction of these scenarios, an attempt was made to cast both tasks in a logical format to expedite the memorization of the task actions.

The go condition description outlined the dominant path for the Data Acquisition task as well as for the Validation task. This condition was comprised of 28 total actions consisting of matrix switches to activate or status field display areas to check.

The no-go and marginal conditions were both based on the dominant path sequence given in the go condition scenario; but also contained corrective or branch actions. The no-go and marginal conditions had a total of 82 and 65 matrix activations or display checks for the operator to memorize.

Two different task scenarios were used in the present study. The first task run was the Data Acquisition task. The subject's task for this portion of the study was to construct and test a printed circuit board (PCB). The major actions for this task required the subject to print circuits, install capacitors and resistors, as well as to test for acceptable current output levels. For the Validation task, the subject's task was to launch, track, command, and receive telemetry from a satellite by manipulating matrix areas and reading status field display areas. Both tasks, construction of the PCBs and satellite tracking, used the same dominant path and branch sequences. The validation portion of this study was primarily used to ensure that the performance scores and errors obtained were due to the HECAD link types, the Teichner translation classifications, or the Information Metric value and not due to scenario differences.

Training session. Training for both the Data Acquisition task and the Validation task required approximately 6 hours. All training was accomplished through the utilization of training tapes and a pictorial mock-up of the actual display. Each session included an explanation of the control (matrix) and display (status field) items, followed by a detailed description or scenario of the specific task to be performed.

Prior to the training session a brief introduction was given to each subject. During this time the mock-up was presented in order to familiarize the subject with the general hardware configuration and the matrix and status field names. In addition to the hardware familiarization, all subjects were given a general overview of the purpose of the experiment, a briefing on the nature of the task, and a true-false quiz on the use of the error keys. This quiz was used to ensure that all subjects were aware of the error situations that went along with either the error feedback or error intuitive keys.

The training sessions for each condition consisted of two repetitions for each segment. For the first repetition, the subject listened to the tape while the experimenter pointed to the corresponding matrix and status field items. This was done in order to familiarize the subject with the correct item locations as well as to eliminate unnecessary scanning of the display by the subject. The phrase "your next action" served as a cue to signify that the next action was being introduced. At the end of a given sequence of actions, the tape was stopped and the subject was asked to repeat the exact sequence of actions while pointing to the correct matrix and status field items. The subject was not allowed to continue to the next action until he had performed the previous one correctly, either with or without the experimenter's assistance.

For both tasks, Acquisition and Validation, the same procedure was used for all three conditions. Time and error scores were recorded during the entire training session; but only for the second repetition of each short sequence of actions. These time and error scores were only used to establish a baseline evaluation for a subject's training performance. If at this time the experimenter estimated that a subject's performance would not reach asymptote in the time available, the training session was terminated and the subject was dismissed. For the Acquisition task, a total of seven out of nineteen individuals were terminated due to poor training performance or cancellations. For the Validation task there were only two cancellations out of ten. Short rest periods, as well as an hour lunch break, were provided throughout the training session.

Practice trials. Three practice trials were given to each subject prior to the actual data collection trials. The practice trials served as a means to acquaint the subject with the actual equipment that they would be using, as well as to give them an example of how the task sequences were put together. During the first two practice trials the experimenter was seated next to the subject in the experimental room. By sitting next to the subject, the experimenter was able to observe how the subject interacted with the equipment, i.e., were they touching the CRT correctly. Also the experimenter could prompt them, if necessary, as to what the next correct control action should be. For the third practice trial, the subject was seated alone in the experimental room and all communications were via an intercom. During this time the experimenter was seated in an adjoining room with the computer system. In addition to being able to observe the subject's performance through a one-way mirror, a second CRT was available for viewing the matrix area and status field. This procedure was continued throughout the entire experimental data collection trials.

Experimental session. The data collection trials required approximately 6 hrs per subject. A total of 37 data collection trials were given to each subject. The subject's task for the Data Acquisition portion of the study was to construct and test a printed circuit board via the touch-entry CRT. The major actions for this task required the subject to print circuits, install capacitors and resistors, as well as to test for acceptable current output levels. For this task a trial consisted of constructing and testing four printed circuit boards for a total of 314 actions. For the Validation portion of the study, the subject's task was to launch, track, command, and receive telemetry from a satellite. As with the Acquisition task, a Validation task trial consisted of tracking and commanding four separate vehicles for a total of 314 actions.

Subjects were not interrupted or prompted during data collection unless they failed to respond correctly within the prescribed period of time. When this time limit of 20 sec was exceeded, the touch entry device would lock up, i.e., no matrix switch inputs were sent to the computer. At this time the experimenter would intervene, tell the subject what the next correct action should be, and re-initiate the computer program to continue the trial.

At the end of every four boards or vehicles the computer system would lock up signifying the end of a trial. At this time the experimenter would ask the subject if he had any comments or problems with the equipment or task. If none, or when the problem was solved, the program was re-initiated and a new trial was started. Rest periods were given at appropriate intervals throughout the data collection trials. Time data, and all matrix switch selections, whether correct or incorrect, were recorded by the computer for all trials.

Preliminary test runs. A total of five pilot subjects were trained and tested on the Data Acquisition task. The data collected were used to validate the training procedures, to establish a time baseline for training and testing, to establish asymptote criteria, and to determine the required number of subjects. The preliminary data collected were not used in the final data analyses for either task. For the Validation task, it was not deemed necessary to train or test pilot subjects because the baseline data collected from the Acquisition task could be generalized to the Validation task.

Debriefing of subjects. Immediately following the last data trial, all subjects were given a questionnaire to fill out. The questionnaire was an attempt to get feedback from the subjects on both the training and data collection portions of the study. Subjects were also asked if they experienced any difficulty learning the task, difficulty using the correct error keys, or problems interacting with the equipment.

SECTION 3

RESULTS

PRE-EXPERIMENTAL ANALYSIS

The following findings were derived before any experimental data were collected for the present study.

Teichner and HECAD Models

Initial efforts were devoted to obtaining a thorough understanding of the theory behind, and implementation of, the Teichner Model. All available documentation of the Teichner translation implementation was concerned with operator task performance during the early stage of the learning curve. Since many of the S-S translations that occur in preasymptotic performance drop out when the operator reaches asymptote, the type and number of S-S translations indicated in this documentation is not wholly applicable to our study contract, where only asymptotic performance is analyzed. Another consideration is the somewhat (admittedly, according to Teichner) subjective technique of applying an S-S translation to a task link. We attempted to solve these difficulties by assuming that only one S-S translation remained for each task link, and then describing in operational terms which type of S-S translation must occur at that task link based upon network configuration.

This assumption is based upon (1) our judgement that compression links will drop out at asymptote and (2) the fact that the S-S creations that appear in Teichner's work are due to multi-use keys, which are not used in our study. A final point in this discussion is the fact that in all cases, in both Teichner's examples and our task network, the S-R translation is the same and therefore a constant.

Examination of the relationships between models has shown that HECAD links and Teichner translations are comparable units of description, even though there is not a high correlation between the values of each. The HECAD link types and Teichner translation types do not track one another in the sense that knowing link type does not specify the type of translation. However, there is a match up between the task descriptors such as link and translation. This match up is shown in Figure 3.

The HECAD task action is shown in the upper half of the figure, and the corresponding Teichner T-task is shown in the lower half. It can be seen that the HECAD stimulus and response nodes correspond with the Teichner stimulus acquisition and R-execution, respectively, and that the HECAD response link matches up with the Teichner S-S and S-R translations. Next to be considered is what happens with error responses. The HECAD link error corresponds with the Teichner translation errors. Both involve a detection stimulus. The HECAD correction link returns to the correct response node, while one or both of the Teichner translations return to the correct R-execution.

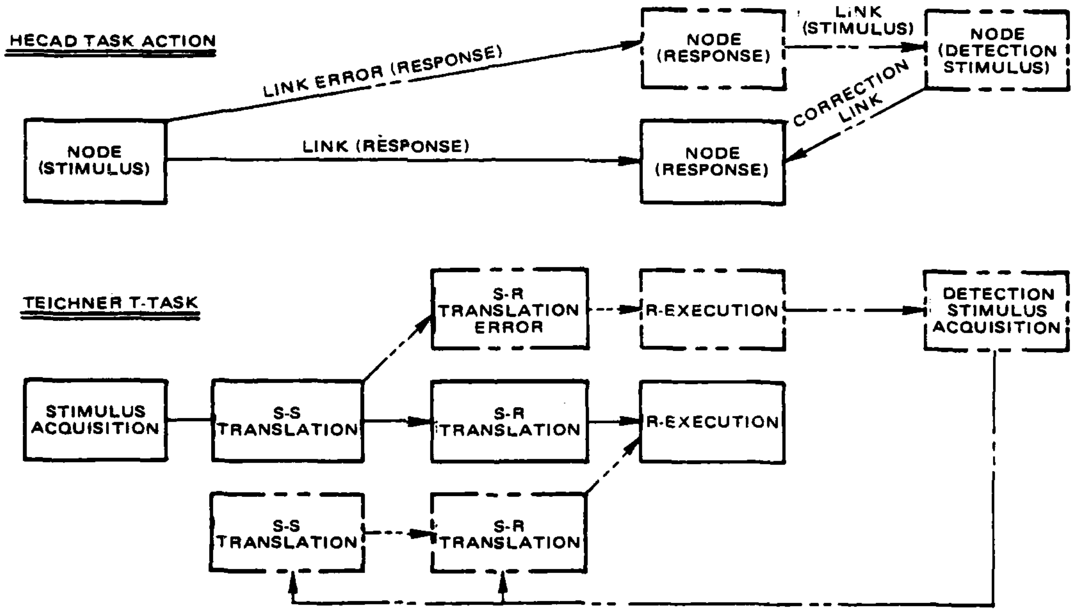


Figure 3. Correspondence of HECAD and TEICHNER Models

Existing Reliability Data

The next step in our work was to determine what, if any, quantitative relationships were shared by the HECAD and Teichner Models, as well as the Information Metric. To accomplish this we used (1) HECAD link type reliability data from Task B of a previous Operator Performance Study (Goldbeck & Charlet, 1975), (2) derived performance reliabilities for each of the three types of Teichner translations (S-S conservation, S-S creation, S-S classification), and (3) derived performance reliabilities for the amount of information processed. The amount of information processed was calculated by the formula:

$$H_c = -\sum^{n} P_i \log_2 P_i$$

Where,

H_c = expression of central processing uncertainty,

n = the number of alternatives,

P_i = the probability of the ith alternative.

The resultant values of this Information Metric were divided into three link types: high information content, medium information content, and low information content, based upon an equal sample size in each group. These three link characteristics were then combined into a 3-dimensional matrix (see Table 1), with reliabilities computed for each cell.

TABLE 1

TASK B RELIABILITIES

INFO METRIC	HECAD LINK TYPE	TEICHNER TRANSLATION TYPE		
		Conservation	Creation	Classification
Low	Recheck	1.0000/4320	--	--
	Dom Path	.9996/5120	--	--
	Mid-branch	.9971/2080	--	--
	Start/return	1.0000/160	--	--
	1-link	---	--	--
Medium	Recheck	---	1.0000/640	1.0000/320
	Dom Path	---	1.0000/2560	.9990/6720
	Mid-branch	---	1.0000/160	.9981/2080
	Start/return	---	.9854/480	.9953/3840
	1-link	---	1.0000/160	.9906/960
High	Recheck	---	1.0000/3040	--
	Dom Path	---	.9981/1600	.9984/1280
	Mid-branch	---	.9988/1600	.9984/640
	Start/return	---	.9938/1440	.9921/2400
	1-link	---	.9979/480	.9292/480

All values given are Reliability/Sample Size

As can be seen from Table 1, only 23 of the possible 45 cells were filled. The unused cells were due to: (1) the division of the Information Metric into three equal samples was such that low information content equated to zero information processed. Creation and classification types of S-S translations, which by definition imply multiple outputs from the initial node of the link, and therefore, information to be processed, cannot occur in a zero information content cell, (2) conversely, cells that contain either medium or high amounts of information processed preclude use of an S-S conservation, since by definition S-S conservations require zero amounts of information, and (3) for the same reason, a HECAD one-link branch cannot be a S-S conservation because a one-link branch, by definition, implies that the initial node has multiple outputs.

Because Task B had been designed without T-Task translations or Information Metric reliabilities in mind, and because of the many resultant low cell frequencies, the results can only be taken as suggestive. Nevertheless, in the nine contrasts of the creation and classification translations, the creation was more or equally reliable than the classification translation in seven of the nine contrasts. Also, in the nine contrasts of the medium amount of information with the high amount, the medium amount of information cells were more or equally reliable than the high information cells in all but two of the contrasts. In the nineteen contrasts of adjacent HECAD link types that were ranked according to reliability expectation, there were only four reversals. Most suggestive of a relationship between the HECAD and Teichner models was the tendency for HECAD link types to be associated with larger performance differences when coupled with a classification than with a creation or conservation. These encouraging results justified the development of tasks that specifically test the integrated model.

Task B marginal reliabilities of the integrated model, giving mean values for the HECAD, Teichner, and Information Metric variables, are shown in Table 2. These values establish the ran^k ordering that is tentatively predicted for the Acquisition task. Results for the Acquisition task will provide a more formal basis of prediction for the Validation task.

Table 2
Task B Marginal Reliabilities

HECAD LINK TYPE	
Recheck	1.0000/8320
Dom Path	.9999/17280
Mid-branch	.9980/6560
Start/return	.9936/8320
1-link	.9788/2080

TEICHNER TRANSLATION TYPE	
Conservation	.9993/11680
Creation	.9982/12160
Classification	.9950/18720

INFORMATION METRIC	
Low	.9993/11680
Medium	.9975/17920
High	.9945/12960

All values given are Reliability/Sample Size

Data Collection Strategy. Error detection and correction prediction algorithms will be examined in a manner similar to performance reliabilities, because we hypothesize that detection and correction probabilities are a function of HECAD link type, T-task translation type, and the amount of information processed. The error detection and error correction probabilities will be computed and correlated with the 3-dimensional cell matrix link-reliability score.

Both the Data Acquisition task and Validation task have been designed using the sample sizes for each cell of the 3-dimensional matrix given in Table 3. Note that with the exception of medium and high information classification rechecks, cells with small sample sizes have been substantially improved from those in Task B. These rechecks were not improved because (1) rechecks have been found to be inherently very reliable and (2) the necessary changes to the tasks to increase these sample sizes would increase task length and, therefore, increase training time, testing time, data reduction and analysis time, with no real payoff.

TABLE 3
DATA ACQUISITION TASK SAMPLE SIZES

INFORMATION METRIC	HECAD LINK TYPE	TEICHNER TRANSLATION TYPE Conservation	Creation	Classif.
Low	Recheck	6000	--	--
	Dom Path	13200	--	--
	Mid-branch	5280	--	--
	Start/return	720	--	--
	1-link	--	--	--
Medium	Recheck	--	4560	480
	Dom Path	--	2880	6960
	Mid-branch	--	1200	1440
	Start/return	--	1920	4800
	1-link	--	1680	1440
High	Recheck	--	5760	--
	Dom Path	--	2880	1440
	Mid-branch	--	3360	1440
	Start/return	--	2880	1440
	1-link	--	2400	1200

EXPERIMENTAL ANALYSES

For all 37 trials, of both tasks, time and error scores were recorded. Time scores consisted of total time to complete each trial as well as individual time scores for each matrix switch area selected. Error scores consisted of a listing of the incorrect matrix switch areas selected. Total number of errors per trial as well as the selected incorrect switch area were also recorded. Also, the number of error feedback and error intuitive key activations were recorded. From these totals, mean values were obtained for every trial (Acquisition task, n=12, Validation task, n=8). Each of these measures will be discussed separately.

Learning/Performance Curves

As can be seen from Figure 4, time and error scores of the Data Acquisition task are relatively high for the beginning trials; however both of these measures decrease rapidly then maintain a steady level. From this figure it was determined that asymptotic performance was reached by trial 13 with occasional time and error fluctuations due to breaks and lunch periods.

In order to further examine the time performance scores, the total time per trial was partitioned into two component parts; namely, the time between two correct responses (C-C) and the time between a correct response and an incorrect response (C-W). Figure 5 illustrates this relationship. As can be seen, the time between two correct actions was very consistent over all trials with little or

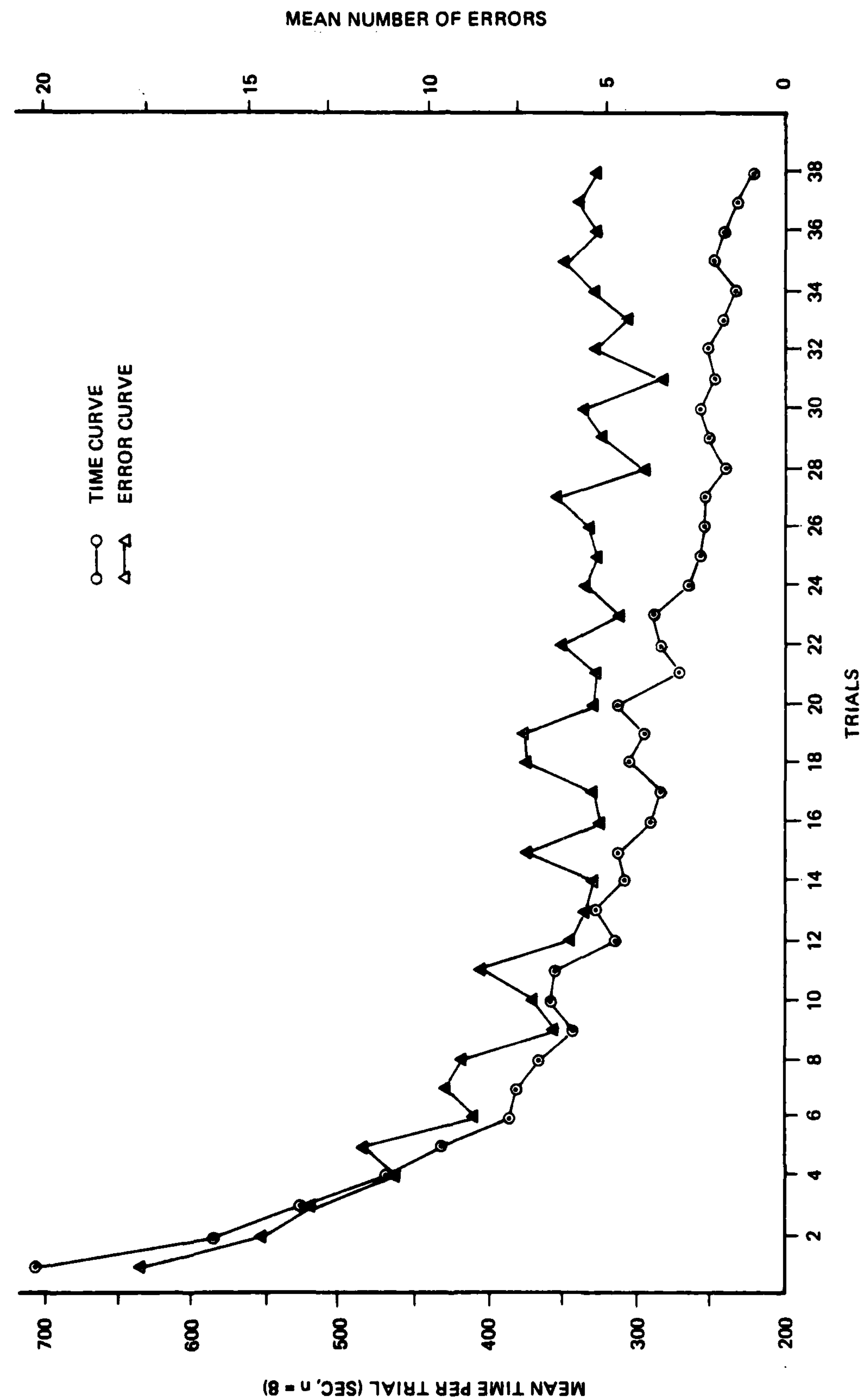


Figure 4. Performance time and errors as a function of trials for the data acquisition task.

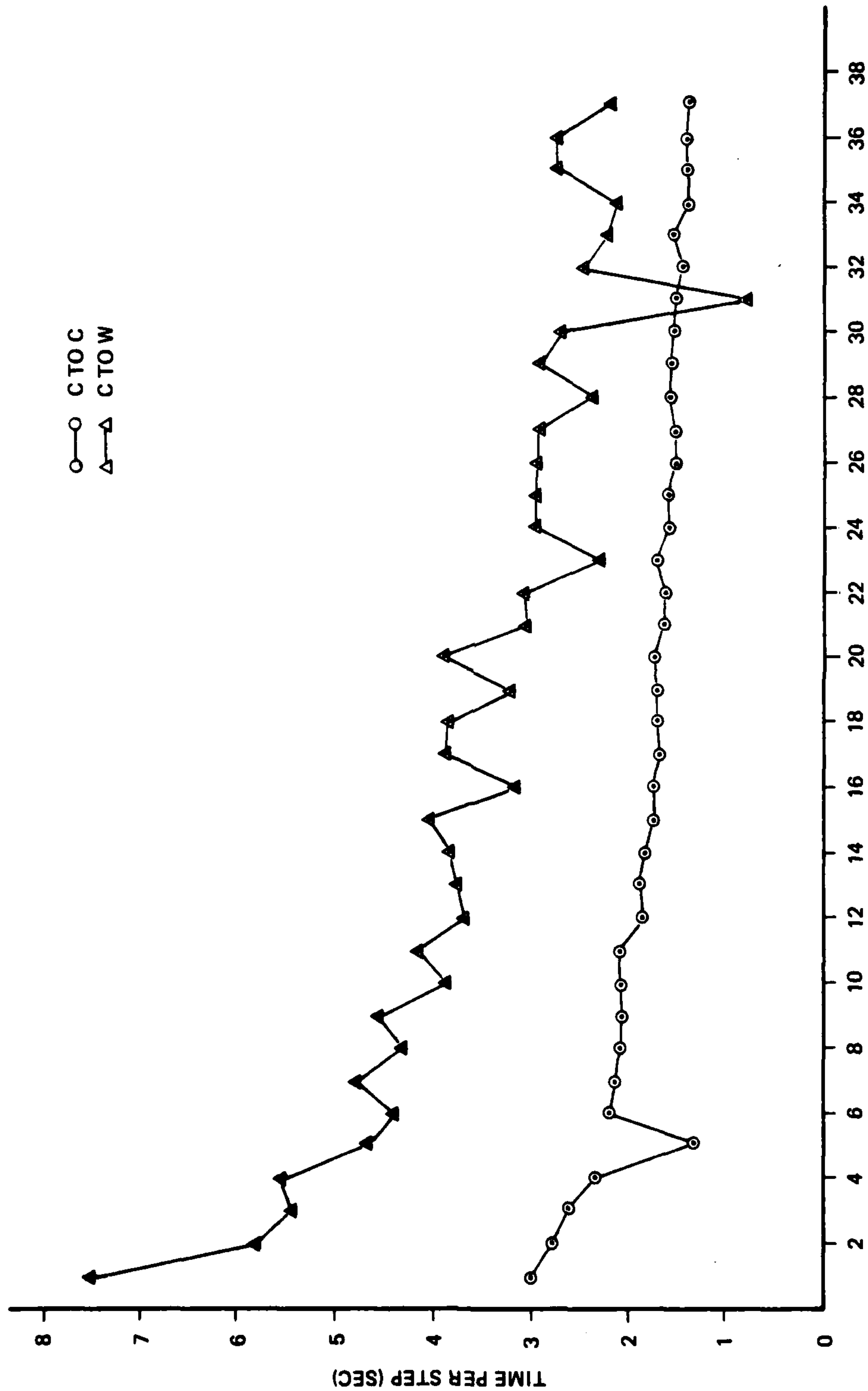


Figure 5. Performance time for correct and incorrect responses as a function of trials for the data acquisition task.

no evident time fluctuations. It is also interesting to note that there is little evidence of learning phenomena. The curve for the C-W responses reflects the amount of time the subject took when making a primary error, i.e., the first incorrect response in any error sequence. This curve clearly shows a classic learning phase followed by asymptotic performance. One would expect that there would be a greater amount of time taken for a C-W response than for a C-C response due to the subject not being able to immediately recall the next correct action to be made. A test of this comparison for asymptotic trials 26-37 confirmed the expectation, $F(1,22)=22.83$, $p < .01$ (C-C response mean = 1.65, C-W response mean = 2.48).

The validation time scores were summarized in the same manner as the acquisition data. Since the tasks had similar values for the integrated model variables, it was predicted that the time score relationships would be similar. Figure 6 depicts the mean number of errors and mean time as a function of trials. As can be seen, there is a decrease in time and error performance over trials, with asymptote being reached by approximately trial 13. In order to further examine the time-error relationships, the C-C responses and the C-W responses were isolated. As with the Acquisition data, it was found that the C-C responses (mean = 1.77) were substantially faster over asymptotic trials than were the C-W responses(mean=2.92), $F(1,22) = 157.80$, $p < .01$. This relationship is illustrated in Figure 7.

An additional analysis was performed for trials 26-37 of the Acquisition task only. This consisted of comparing the response time for

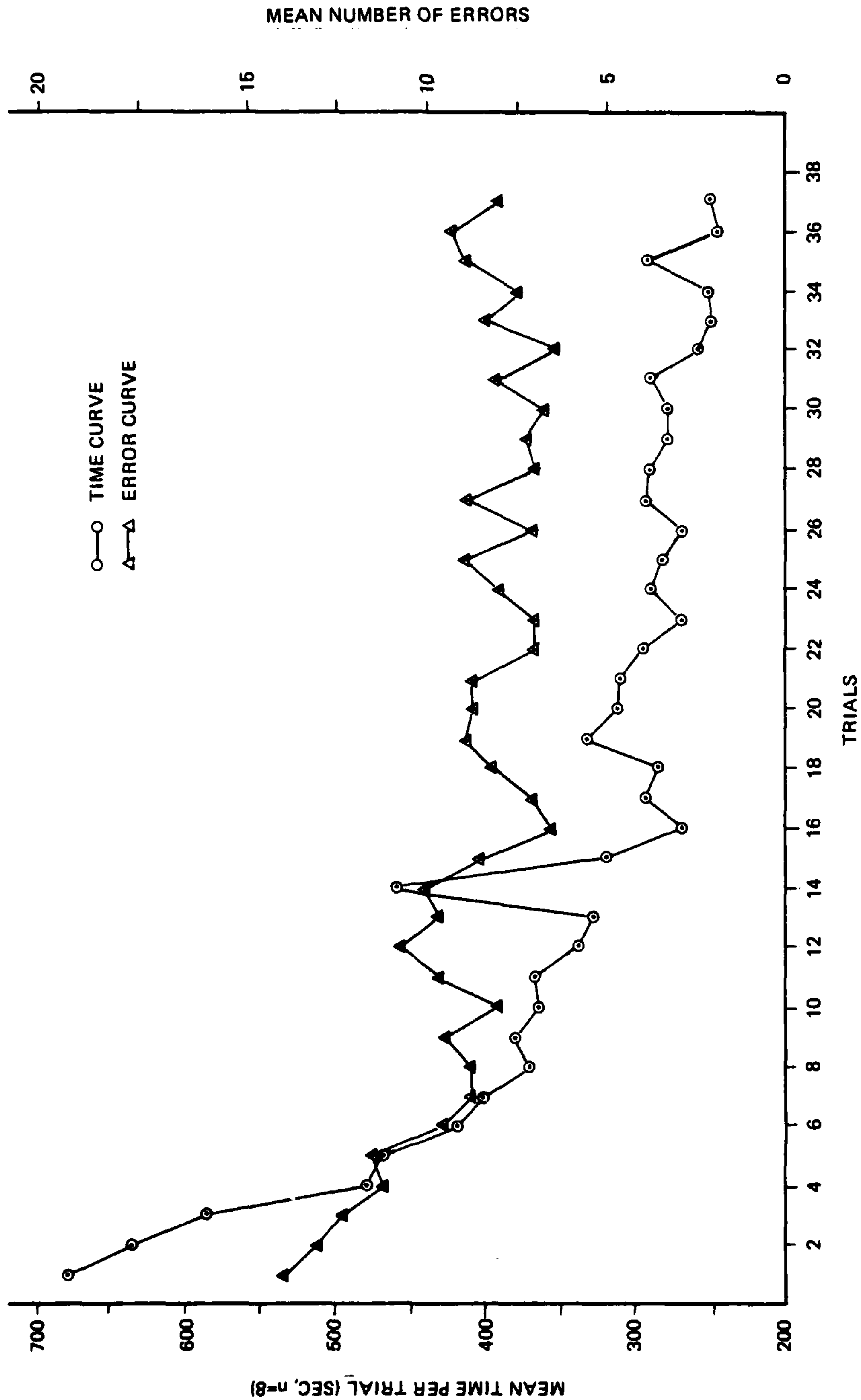


Figure 6. Performance time and errors as a function of trials for the validation task.

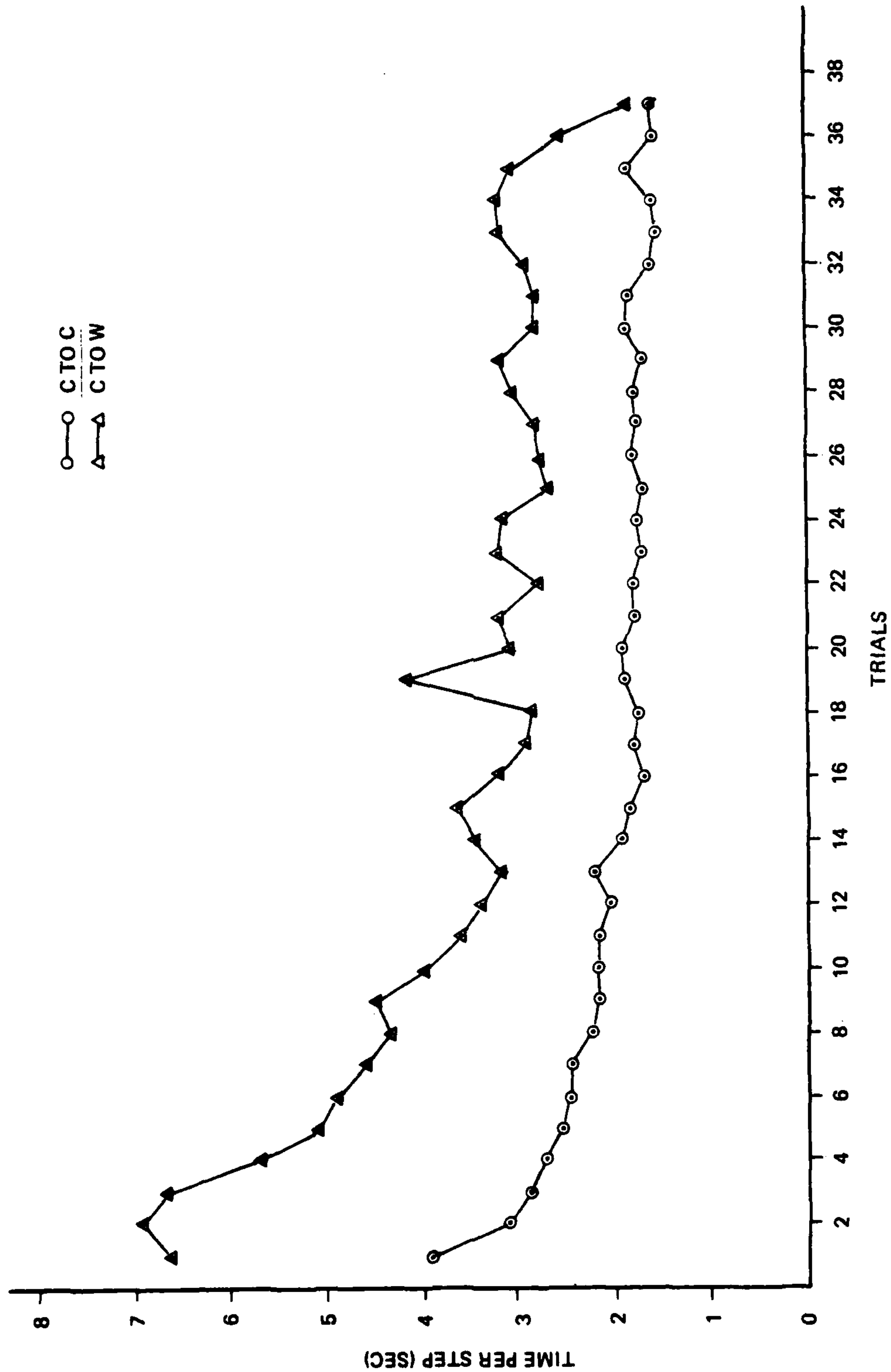


Figure 7. Performance time for correct and incorrect responses as a function of trials for the validation task.

the first to second incorrect response with the response times among all incorrect strings of responses. This comparison reached a conventional level of significance, $F\ (1, 22) = 15.57$, $p < .01$ with the time between the first two incorrect responses being less than the time between subsequent incorrect responses. A reasonable explanation for this finding concerns the procedure for signaling subjects that an error had been made. This signal came only after a string of two errors had been made. It is suggested that on many occasions the subjects were unaware of having made an error until the signal came on, and hence performed the second error in the string as rapidly as if no error had been made. For the comparison of correct-correct responses with incorrect #1-incorrect #2 responses there was no difference ($F < 1$), with the correct-correct responses being slightly faster. However, once the error signal came on, response performance was slowed by the cognitive processes of attempting error correction.

Error Detection and Correction Time Scores

As described in the Method section, two matrix switches were dedicated to signaling a feedback detection and an intuitive detection. It was initially hypothesized that the detection of a primary error detected intuitively would take less time than an error detected through feedback. Figures 8 and 9 show a performance tendency that favors this hypothesis during preasymptotic trials, but not for trials after asymptote was reached. There were no significant differences on tests run for asymptotic trials 26-37, so statistical tests were run using all trials. For the Acquisition task, it was found that

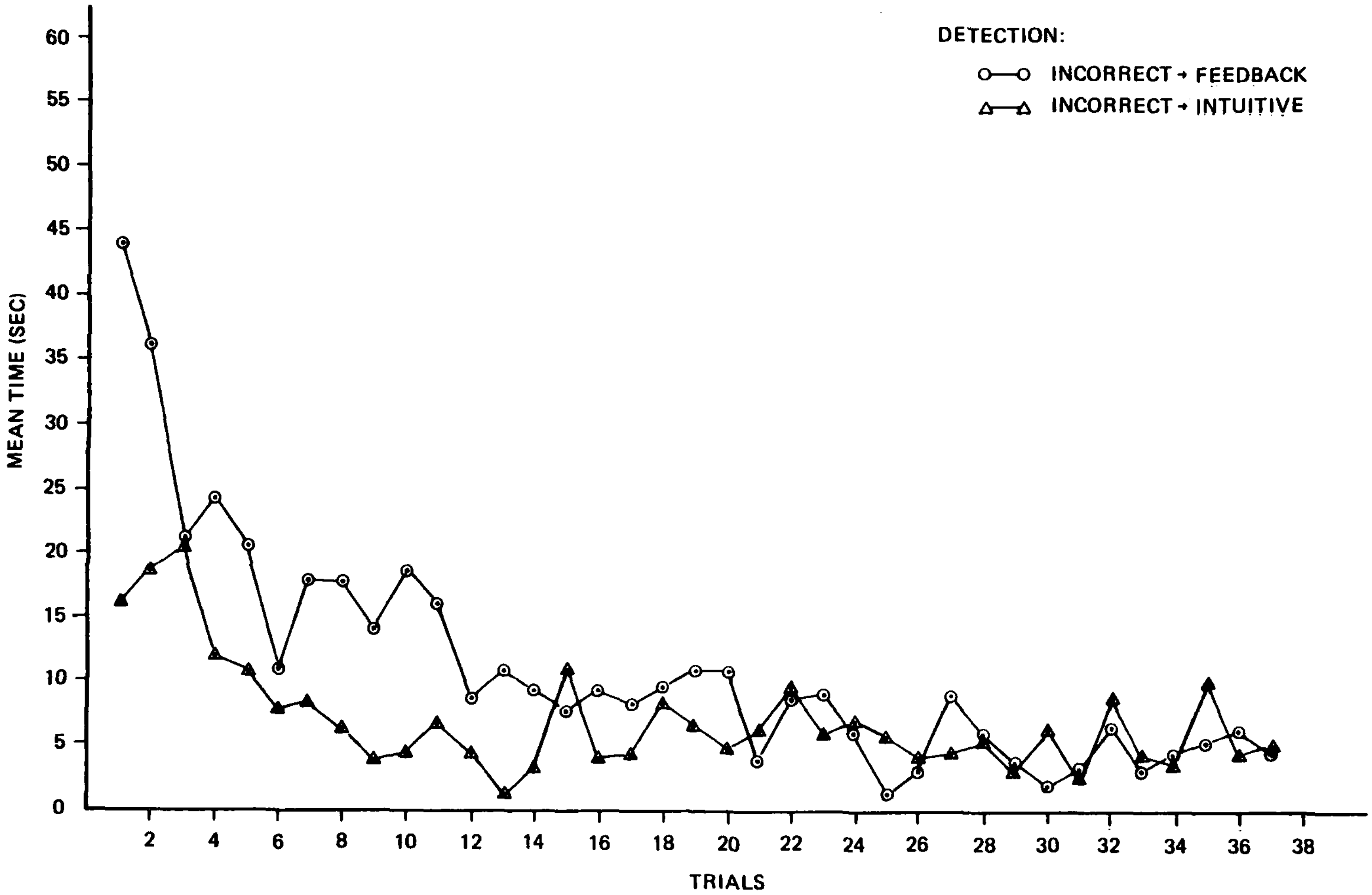


Figure 8. Time to a feedback detection and an intuitive detection as a function of trials for the data acquisition task.

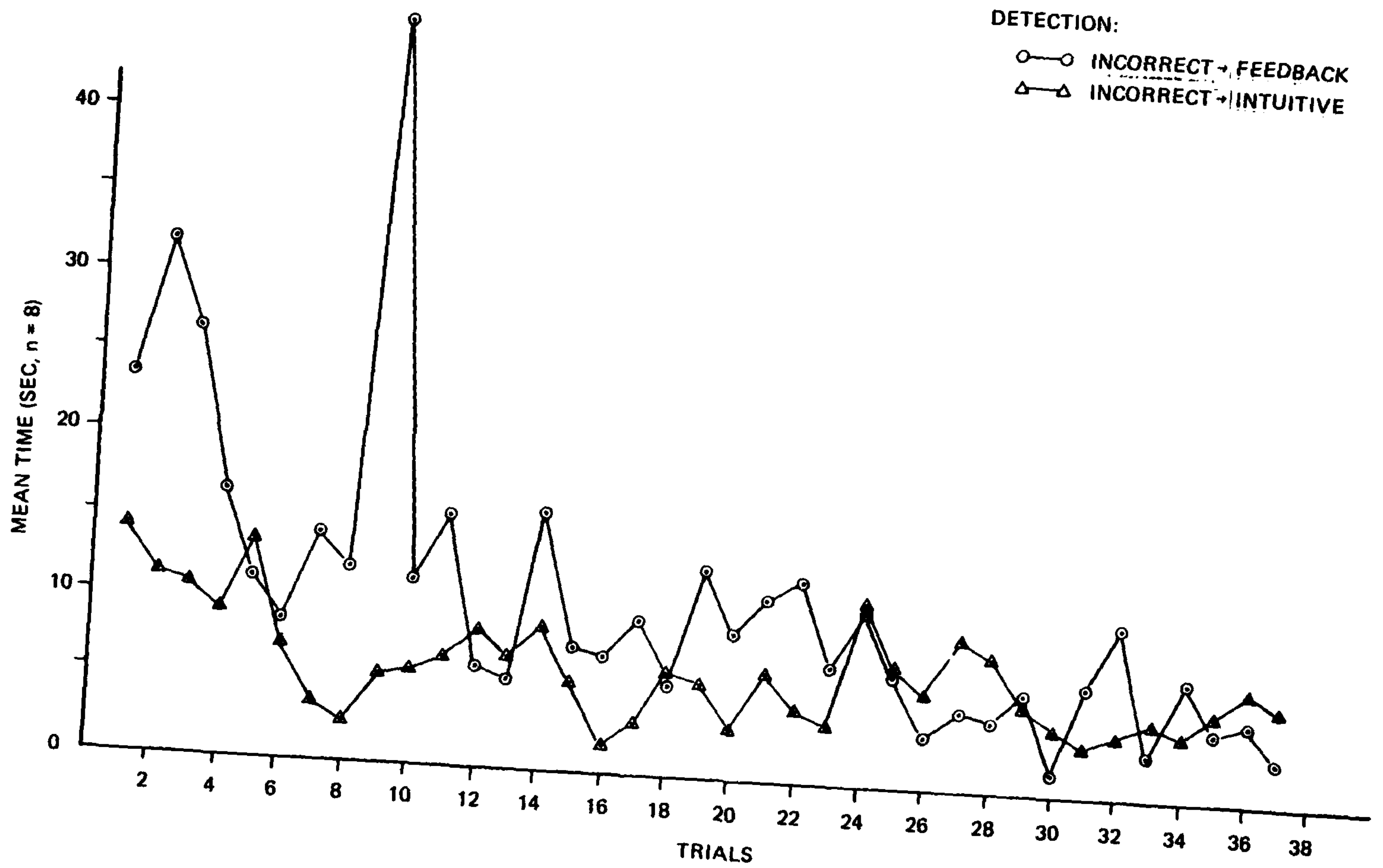


Figure 9. Time to a feedback detection and an intuitive detection as a function of trials for the validation task.

there was a significant difference, $F(1,72) = 6.84$, $p < .05$, between the incorrect response followed by a feedback detection and an incorrect response followed by an intuitive detection. This result was in the predicted direction as it took longer (approximately 4 sec) to signal that an error was made when the subject had to rely on feedback information. This finding was substantiated on the Validation task, $F(1,72) = 5.52$, $p < .05$, with feedback detected errors requiring more time than intuitively detected errors. These findings indicate that the processes involved in detecting an error by means of feedback are complicated enough to consume a substantial amount of time, but are of a nature that is susceptible to a learning that reduces the time to a period comparable to the intuitive process.

For the correction portion of an error, it was initially hypothesized that it would take a greater amount of time to correct an error that was preceeded by the feedback detection key than the intuitive detection key. There were no significant differences on tests run over asymptotic trials 26-37. While the results for overall trials on the Acquisition task are in the correct direction, significance was not obtained, $F(1,72) = 3.80$, $p > .05$. For the Validation task, however, significance was reached, $F(1,72) = 5.65$, $p < .05$, with feedback detected errors requiring substantially more time than intuitively detected errors prior to reaching the next correct response. Both results are depicted graphically in Figures 10 and 11. For both tasks the time differences are largely attributable to the subjects' preasymptotic performance. Once asymptote was reached, the differences in time between intuitive and feedback for correction of an error were reduced. Explanation of this "cor-

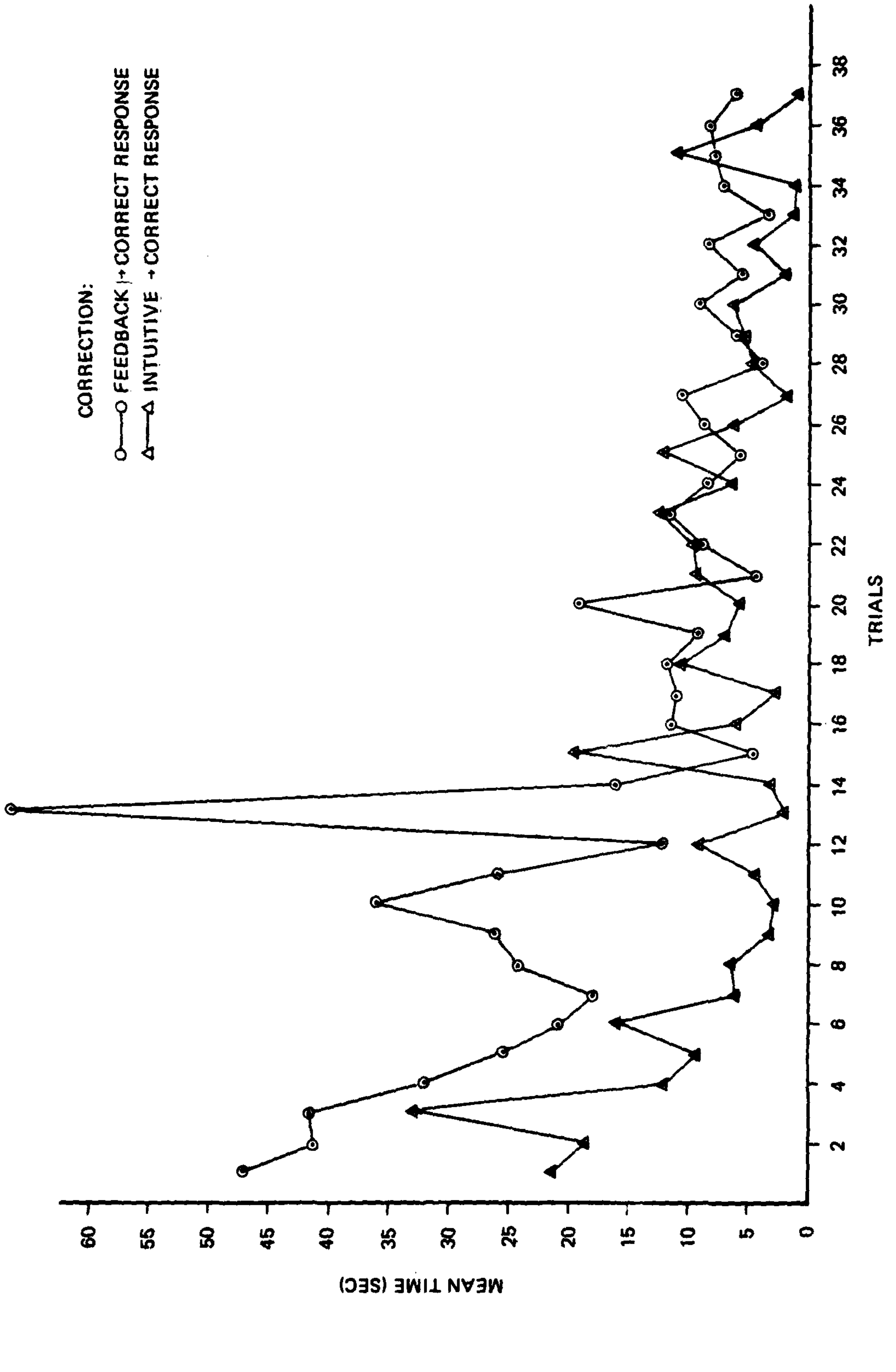


Figure 10. Time to correction following a feedback detection and an intuitive detection as a function of trials for the Data Acquisition task.

Ford Aerospace &
Communications Corporation

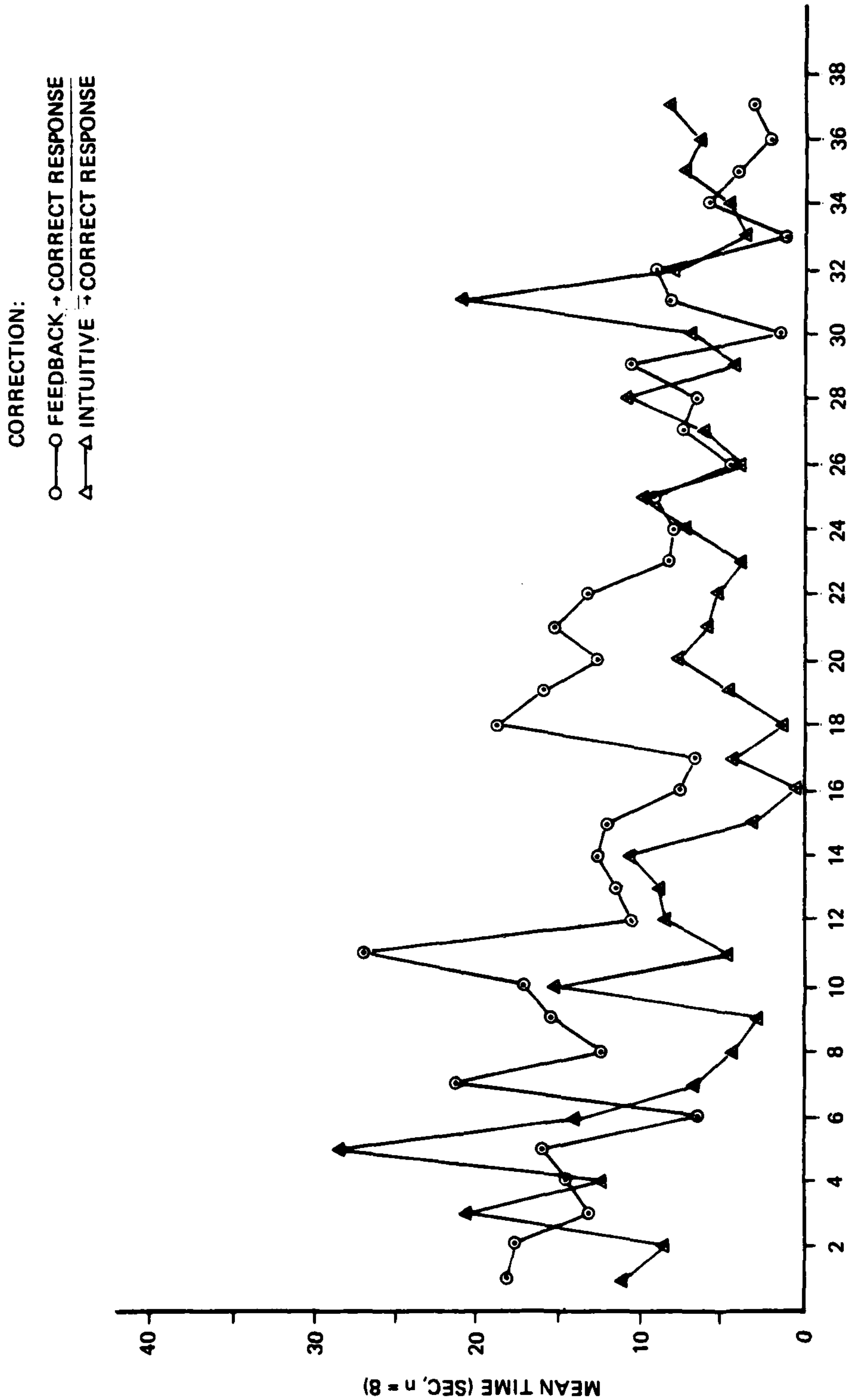


Figure 11. Time to correction following a feedback detection and an intuitive detection as a function of trials for the validation task.

rection" finding is similar to the explanation advanced for the "detection" finding, in that both represent an early trial increment of processing time, associated with detection via feedback, that dissipates with learning trials. What is not clear from the data is whether the subjects became more efficient in their capability to scan and locate display items, or to process cognitive material, or both.

Related Findings

According to Rabbitt (1969) the response time to the first event after an error is very slow, with the responses to the next few events also being slower than average. Furthermore, Rabbitt (1966a, 1966b, 1968) found that errors were much faster than correct responses. Subjects make runs of increasingly fast responses, which terminate in an even faster error. Evidence for none of these findings was found in the present study which in fact showed opposite effects. Rabbitt used a choice reaction time task, while the present study had a stronger cognitive task element with a heavy dependence on long term memory. It is apparently this task type difference that accounts for the differences in times associated with error events.

Node vs Link Errors

Findings from previous studies (Goldbeck & Charlet, 1975) have shown that task time and reliability are determined by task structure. In that study it was shown that task structure characteristics can be expressed in terms of parameters describing how a system task at a control-display panel is performed differently on different occasions. In order to develop possible prediction algorithms for operator performance, as well as error detection and correction, the task must be represented as a network of control and display links. Before presenting the actual results, a brief summary will be devoted to the models for operator performance and how the actual values for the control and display links are derived.

Performance models. Three different models were combined in an attempt to predict operator task times and error rates. The first model was developed by AMRL in cooperation with Ford Aerospace, and is referred to as HECAD (Human Engineering Computer Aided Design). This model assigns values to either control or display links. From each control or display (node) there are one or more links to other nodes. For each link there are different associated probabilities that an action will occur. The path of highest link probablity through the task network is the dominant path. When performance of the task departs from the dominant path of high probability links, these departures are referred to as branch paths. In addition to the dominant path links, there are: 1) the first and last links of a branch, or start-returns, 2) the middle links of a branch, each called a mid-branch, 3) 1-link branches, i.e; a non-sequential jump from one dominant path node to another, and 4) rechecks. The recheck link category is not mutually exclusive with respect to the other four types but this designation supercedes all other categories. All link errors were assigned to the appropriate link type.

The second model for performance was developed by Warren Teichner. The three corresponding features for this model are: 1) conservation, signifies a one to one relationship, 2) classification, a many to few relationship, and 3) creation, a few to many relationship. All link errors were assigned to the appropriate Teichner translation type.

The third model was an Information Metric value based upon the formula $-\sum P_i \log_2 P_i$. The resultant values were divided into three link types: 1) high information content, 2) medium information content, and 3) low information content. These three link characteristics were then combined into a 3-dimensional matrix with reliabilities computed for each cell.

Link-node assignments. All asymptotic trial errors were divided into two types, namely, link errors and node errors. Node errors result when errors of manipulation occur at the correct or adjacent control or display, i.e., the operator was at the correct point in the task conceptually, but did not sucessfully perform the correct action. Link errors were the result of an incorrect cognitive transfer from the last node he sucessfully performed to the next node.

Once an error was assigned to a link or node it was then assigned to a control or display. Display errors were categorized by the following parameters: 1) control is correct for different status of display, 2) control is not correct for any state of preceeding display, 3) control is correct for an adjacent display of the same type and status, and 4) time limit: no control is selected. Similarly, errors were categorized as control errors when any one of the following criteria was met: 1) similar control names, but an incorrect or adjacent control is touched, 2) control preceded by another control with no intervening display, 3) similar control name and in vicinity of correct control, and 4) time limit: a series of incorrect controls are selected.

Overall reliability results. For both tasks all link errors were categorized into a three dimensional matrix using the HECAD, Teichner, and Information Metric values. From this matrix an overall link error rate table was derived. Proportions were calculated, based on the total number of link errors per cell, corrected for the cell frequency, number of subjects, and trials. Lastly, a reliability measure was calculated for each cell.

As expected from Task B results of the previous study, HECAD link type summaries of asymptotic trials from the Data Acquisition task in Table 4 shows the same order from most to least reliable links: re-check, dominant path, mid-branch, start/return, and 1-link. The rationale for these findings with the empirically derived categories includes the fact that rechecks should have been easy because they were strongly cued by an identical response. Dominant path links should have been easy due to their high probability of occurrence and the fact that they were the nominal response in the path of task resonses. Mid-branches were moderately well cued by preceding branch links, but start/returns represented a shift to or from a branch. The 1-link was a low probability event that revised normal dominant path relationships. A parallel rationale for the results concern

Table 4

Overall Link Reliability Rates: Acquisition Task

INFO METRIC	HECAD LINK TYPE	TEICHNER TRANSLATION TYPE Conservation	Creation	Class	SUMMARY
Low	Recheck	1.0000	--	--	.9946
	Dom Path	.9915	--	--	
	Mid-branch	.9965	--	--	
	Start/return	.9907	--	--	
	1-link	--	--	--	
Medium	Recheck	--	.9901	.9545	.9845
	Dom Path	--	.9873	.9890	
	Mid-branch	--	.9911	.9889	
	Start/return	--	.9705	.9740	
	1-link	--	.9708	.9479	
High	Recheck	--	1.0000	--	.9746
	Dom Path	--	.9878	.9919	
	Mid-branch	--	.9787	.9444	
	Start/return	--	.9311	.9734	
	1-link	--	.9576	.9639	
SUMMARY		.9946	.9811	.9783	

HECAD LINK TYPE SUMMARIES

Recheck	.9960
Dom Path	.9901
Mid-branch	.9951
Start/return	.9667
1-link	.9612

the general correspondence of response reliability with the relative amount of practice within the task. The Validation task results for HECAD link types shown in Table 5 have the predicted order of reliability with the exception of a reversal between start/return and 1-link.

For both tasks, the low information metric category resulted in the highest reliability value and the high information metric category resulted in the lowest reliability value. This finding is in keeping with the expectation that higher uncertainty will be associated with more errors.

For the Teichner classifications it was found that the conservation, that is, a node with one state and one output, yielded a higher reliability than the creation or classification for both tasks. A surprising finding is that the creation translation was more reliable than classification translation for both tasks, as well as for the previous Task B. It would have been more plausible if the creation had been least reliable because multiple outputs are possible for it, depending on the status of a previous node. For a classification there is only one output possible for a multi-state node.

Despite the similarity between the Data Acquisition and Validation tasks in regard to their having the same dominant path and branch sequences, their overall reliability differed substantially. The Data Acquisition task had a reliability of .9850 and the Validation task had a .9775 reliability. It is speculated that the most likely area for finding parameters that would account for a major part of

Table 5

Overall Link Reliability Rates: Validation Task

INFO METRIC	HECAD LINK TYPE	TEICHNER TRANSLATION TYPE Conservation	Creation	Class	SUMMARY
Low	Recheck	.9987	--	--	.9953
	Dom Path	.9962	--	--	
	Mid-branch	.9915	--	--	
	Start/return	.9792	--	--	
	1-link	--			
Medium	Recheck	--	1.0000	1.0000	.9720
	Dom Path	--	.9818	.9810	
	Mid-branch	--	.9524	.9927	
	Start/return	--	.9661	.9292	
	1-link	--	.9485	.9740	
High	Recheck	--	1.0000	--	.9645
	Dom Path	--	.9627	.9983	
	Mid-branch	--	.9754	.9410	
	Start/return	--	.9167	.9549	
	1-link	--	.9135	.9833	
SUMMARY		.9953	.9695	.9833	

HECAD LINK TYPE SUMMARIES

Recheck	.9996
Dom Path	.9874
Mid-branch	.9781
Start/return	.9384
1-link	.9422

this difference is the area of task scenario, or story line. Such characteristics as simplicity and redundancy could be scaled.

Detection reliabilities. Tables 6 and 7 show the detection proportions for intuitively detecting an error. These proportions were calculated by taking the total number of intuitive detections and dividing this total by the number of link errors.

For both the Acquisition and Validation tasks, the high categorization under the Information Metric resulted in the highest reliability. For the medium and low information categories there was no consistency between tasks; with the low information metric having a higher reliability than the medium for the Acquisition task and the converse situation for the Validation task.

The Teichner classifications did not reveal any predictive or consistent measures between the two tasks. The Validation task shows that a link error that occurred on a node with multiple outputs was more likely to be detected than a node error where there was only one output. This is also upheld for the Acquisition task as the creation category shows a higher detection rate than the classification category. However, unlike the Validation task, the conservation category yielded an even higher error reliability figure than both the creation and classification categories, thus making interpretation difficult. This increase in intuitive detection of an error could be attributable to the fact that a subject had several sources of information available to him in order to cue him that something had gone awry in the operating system. Similarly, if the subject had only

Table 6

Proportions for the Intuitive Detection
of Primary Error: Acquisition Task

INFO METRIC	HECAD LINK TYPE	TEICHNER TRANSLATION TYPE Conservation	Creation	Class	SUMMARY
Low	Recheck	.0000	--	--	.2821
	Dom Path	.2687	--	--	
	Mid-branch	.3636	--	--	
	Start/return	.0000	--	--	
	1-link	--	--	--	
Medium	Recheck	--	.0000	.4616	.4784
	Dom Path	--	.5909	.6522	
	Mid-branch	--	.3333	.6250	
	Start/return	--	.6364	.4533	
	1-link	--	.6667	.3333	
High	Recheck	--	.0000	--	.2615
	Dom Path	--	.4762	.2857	
	Mid-branch	--	.4419	.1250	
	Start/return	--	.2017	.2074	
	1-link	--	.3115	.2308	
SUMMARY		.2821	.3263	.3801	

HECAD LINK TYPE SUMMARIES

Recheck	.1500
Dom-Path	.4479
Mid-branch	.3307
Start/return	.3017
1-link	.3577

Table 7

Proportions for the Intuitive Detection of a Primary Error: Validation Task

INFO METRIC	HECAD LINK	TEICHNER TRANSLATION TYPE Conservation	Creation	Class	SUMMARY
Low	Recheck	.6667	--	--	.4861
	Dom Path	.4500	--	--	
	Mid-branch	.5000	--	--	
	Start/return	.3333	--	--	
	1-link	--	--	--	
Medium	Recheck	--	.0000	.0000	.2810
	Dom Path	--	.5238	.1698	
	Mid-branch	--	.0000	.0000	
	Start/return	--	.5769	.3088	
	1-link	--	.2692	.4000	
High	Recheck	--	.0000	--	.1883
	Dom Path	--	.1628	.0000	
	Mid-branch	--	.4848	.1471	
	Start/return	--	.1354	.3077	
	1-link	--	.1084	.3750	
SUMMARY		.4681	.2167	.2556	

HECAD LINK TYPE SUMMARIES

Recheck	.6667
Dom Path	.2609
Mid-branch	.2419
Start/return	.2759
1-link	.1721

one clue in the operating system that an error had occurred, this clue may be overlooked, thereby lowering the detection reliability.

As with the Tiechner classifications, HECAD reliability figures were not consistent between tasks. Due to this inconsistency it is difficult to draw any substantial conclusions regarding this variable (see Tables 6 and 7).

In addition to just detecting an error there is also detection immediately followed by a correct response. Tables 8 and 9 summarize the proportions for the intuitive detection of a primary error resulting in a correct response. This aspect will be briefly discussed by its impact on Teichner, HECAD, and the Information Metric categories.

Results show that under the Teichner categories the classification (many to few) grouping yielded the largest proportion of detections with correction for both tasks. While this finding is consistent between tasks, the conservation and classification categories were in opposite directions. Due to this inconsistency, interpretation is difficult with the amount of information we presently have.

For the HECAD categories there was no consistency between the two tasks. Due to this inconsistency, no conclusions or speculations will be made regarding this variable. The obtained results for the Information Metric were similar, but only for the medium information category. For both tasks, the medium category resulted in a high proportion of errors being detected and corrected. For

Table 8

Proportions for Intuitive Detection of a Primary Error Resulting in a Correct Response: Acquisition Task

INFO METRIC	HECAD LINK TYPE	TEICHNER TRANSLATION TYPE Conservation	Creation	Class	SUMMARY
Low	Recheck	.0000	--	--	.6923
	Dom Path	.7778	--	--	
	Mid-branch	1.0000	--	--	
	Start/return	.0000	--	--	
	1-link	--	--	--	
Medium	Recheck	--	.0000	.0000	.9344
	Dom Path	--	1.0000	1.0000	
	Mid-branch	--	1.0000	1.0000	
	Start/return	--	1.0000	1.0000	
	1-link	--	.7143	.8000	
High	Recheck	--	.0000	--	.7692
	Dom Path	--	1.0000	.5000	
	Mid-branch	--	.5789	.6667	
	Start/return	--	.7500	1.0000	
	1-link	--	.7895	1.0000	
SUMMARY		.6923	.7982	.9327	

HECAD LINK TYPE SUMMARY

Recheck	1.0000
Dom Path	.9315
Mid-branch	.7619
Start/return	.8243
1-link	.7955

Table 9

Proportions for Intuitive Detection of a Primary Error Resulting in a Correct Response: Validation Task

INFO METRIC	HECAD LINK	TEICHNER TRANSLATION TYPE Conservation	Creation	Class	SUMMARY
Low	Recheck	.5000	--	--	.8636
	Dom Path	.8889	--	--	
	Mid-branch	.8889	--	--	
	Start/return	1.0000	--	--	
	1-link	---			
Medium	Recheck	--	.0000	.0000	.9651
	Dom Path	--	.9091	.8889	
	Mid-branch	--	.0000	.0000	
	Start/return	--	.9333	1.0000	
	1-link	--	1.0000	1.0000	
High	Recheck	--	.0000	--	.7705
	Dom Path	--	.8571	.0000	
	Mid-branch	--	.5625	.8000	
	Start/return	--	1.0000	.7500	
	1-link	--	.6667	1.0000	
SUMMARY		.8636	.8333	.9420	

HECAD LINK TYPE SUMMARY

Recheck	.5000
Dom Path	.8889
Mid-branch	.7000
Start/return	.9625
1-link	.8571

the low and high information categories there was no consistency between tasks; with the high information metric having a higher detection and correction proportion than the low information metric for the Acquisition task and the converse situation for the Validation task.

Feedback detected errors. As mentioned previously, a feedback detected error refers to the fact that the subject required some outside or additional information to inform him that an error had been made. Tables 10 and 11 summarize these findings. For the Information Metric variable there was no consistency between tasks for the proportions or reliabilities of detected errors. One point that should be made however, is that there are relatively few errors that come under the feedback detected error category (i.e., 17 out of 677); therefore it is extremely difficult to draw any definitive conclusions. For the Teichner categories, both tasks resulted in the M2 classification (many to few) yielding the highest reliability. An examination of Tables 10 and 11 show that caution should be taken with this interpretation as it is based on only six observations, thus making interpretations of the data difficult to make. The HECAD categories yielded little insight into the feedback detection question. Again, this inconsistency between tasks is basically attributable to the small number of errors detected through feedback.

In addition to detection of an error with the feedback key, it was decided to examine the correction rate immediately following the feedback error key. The proportion of errors that were initially activation of the detected and feedback corrected was high for both

Table 10

Proportions for the Feedback Detection of a Primary Error: Acquisition Task

INFO METRIC	HECAD LINK TYPE	TEICHNER TRANSLATION TYPE Conservation	Creation	Class	SUMMARY
Low	Recheck	.0068	--	--	.0175
	Dom Path	.0485	--	--	
	Mid-branch	.0000	--	--	
	Start/return	.0000	--	--	
	1-link	--	--	--	
Medium	Recheck	--	.0631	.0000	.0497
	Dom Path	--	.0000	.0000	
	Mid-branch	--	.0000	.0000	
	Start/return	--	.5000	.0000	
	1-link	--	.0000	.0000	
High	Recheck	--	.0435	--	.0192
	Dom Path	--	.0000	.0000	
	Mid-branch	--	.0000	.0000	
	Start/return	--	.0000	.0000	
	1-link	--	.0000	.0000	
SUMMARY		.0175	.0495	.0000	

HECAD LINK TYPE SUMMARY

Recheck	.0335
Dom Path	.0485
Mid-branch	.0000
Start/return	.0108
1-link	.0000

Table 11

Proportions for the Feedback Detection
of a Primary Error: Validation Task

INFO METRIC	HECAD LINK TYPE	TEICHNER TRANSLATION TYPE Conservation	Creation	Class	SUMMARY
	Recheck	.0000	--	--	
	Dom Path	.0500	--	--	
Low	Mid-branch	.0000	--	--	.0362
	Start/return	.1667	--	--	
	1-link	--	--	--	
	Recheck	--	.0000	.0000	
	Dom Path	--	.0476	.0377	
Medium	Mid-branch	--	.0313	.0000	.0327
	Start/return	--	.0000	.0221	
	1-link	--	.1154	.0000	
	Recheck	--	.0000	--	
	Dom Path	--	.0233	.0000	
High	Mid-branch	--	.0303	.0294	.0154
	Start/return	--	.0208	.0000	
	1-link	--	.0000	.0000	
SUMMARY		.0362	.0250	.0222	

HECAD LINK TYPE SUMMARY

Recheck	.0000
Dom Path	.0362
Mid-branch	.0242
Start/return	.0207
1-link	.0246

tasks (mean = .94). Due to the limited number of cases however, one must be judicious in the interpretation of these data.

NODE ERRORS

As defined previously, node errors are those errors which can be attributed to a mismanipulation of the matrix area. For the Acquisition task, post-asymptotic trials only, reliability figures were obtained for the three variables; namely, the Information Metric, Teichner classifications, and HECAD categories. It was originally hypothesized that the node error rate would not lend itself to be a viable predictor variable for operator performance, but would prove to be an overall baseline error rate. This baseline error rate would be representative of those errors that would occur during operating periods due to mismanipulations of the equipment, lapses of attention (i.e., time out errors), or inadvertently activating an adjacent switch area.

An examination of Table 12 shows that for the HECAD categories there was no significant difference between the types of links. Therefore, this baseline error rate and the reliability estimates are monotonic with respect to the varying levels and difficulty of the HECAD categories. This finding also extends to the Information Metric and the Teichner categories. An inspection of the overall reliabilities shows that the occurrence of a node error was likely but at a low level. Additionally, they were normally distributed throughout the task, i.e., there was no particular bias by category (Acquisition task reliability = .9978, Validation task

Table 12

Node Reliability Values for the
Data Acquisition Task

INFO METRIC	HECAD LINK	TEICHNER TRANSLATION TYPE Conservation	Creation	Class	SUMMARY
Low	Recheck	1.0000	--	--	.9989
	Dom Path	.9995	--	--	
	Mid-branch	.9981	--	--	
	Start/return	.9970	--	--	
	1-link				
Medium	Recheck	--	1.0000	1.0000	.9965
	Dom Path	--	.9931	.9943	
	Mid-branch	--	.9875	1.0000	
	Start/return	--	.9939	.9983	
	1-link	--	.9875	1.0000	
High	Recheck	--	1.0000	--	.9982
	Dom Path	--	.9948	.9977	
	Mid-branch	--	.9985	.9977	
	Start/return	--	.9959	1.0000	
	1-link	--	.9948	.9977	
SUMMARY		.9989	.9973	.9972	

HECAD LINK TYPE SUMMARY

Recheck	1.0000
Dom Path	.9969
Mid-branch	.9981
Start/return	.9970
1-link	.9968

reliability = .9982).

ADDITIONAL ANALYSES

It was originally hypothesized that detection and correction probabilities were a function of HECAD link types, Teichner translation types, and the amount of information processed. Like performance reliabilities, error detection and correction data could be examined in a similar manner. In the present study however, the error sample size is a subset of the total for a given cell in the three dimensional matrix. The resultant error detection and correction sample size was insufficient to determine necessary prediction algorithms. Therefore, the technique of comparing the predicted with the actual value for each cell was not successfully used to examine detection and correction probabilities. Instead, the error detection and correction probabilities were computed and plotted. Using scatter plots, correlation coefficients were calculated using the cell performance reliability and the probability of detection of an error on the Data Acquisition task, as well as the Validation task. This design yielded six correlations per task based on the Information Metric categories.

For the Data Acquisition task, the low information category yielded the largest correlation coefficient (r=+79). Before drawing any parallels regarding the overall error and the detection probabilities, one should look at the other correlation coefficient values for the remaining information categories to see if any patterns are evident. For the medium information category a correlation of +.08 was obtained;

for the high information category the correlation value was +.67. Corresponding correlations were run for the correction data, but they were negative.

Clearly there is no simple interpretation of the data. One possible explanation for the low information category yielding a high correlation is the assumption that it was very easy for a subject to detect that an error had occurred. Therefore, the higher the reliability rate, the greater the chance for detection, and since the errors are readily apparent to the operator there would be a strong relationship for errors being detected. This explanation, however, does not fully explain the obtained results; therefore, an examination of the validation correlation coefficients was necessary. The low information category on this task yielded a high correlation (+.84), while the correlations for the medium and high category were +.25 and +.26 respectively. The additional correlations when considered separately did not lend further insight, so an overall correlation coefficient was calculated for each task. For the Acqusition task this yielded a value of +.51, for the Validation task a value of +.45 was obtained. While this additional information does not provide a solution, it does show that the degree of relationship between the two variables is similar, and substantial for both tasks.

The following simple algorithm is a convenient means for showing the effects of detection and correction on task reliability:

$$R_A = R_a + [(1-R_a)(D_a)(C_a)]$$

Where R_A = operator reliability for task action A, i.e., the probability that task action A will be correctly performed (including error detection and correction).

R_a = The probability that task action A will be performed without error.

D_a = The probability that the operator detects an error in the performance of task action A.

C_a = The probability that a detected error is successfully corrected by the operator.

While in the present study there was not found any prediction functions for the values of D or C, some information can be gained by presenting the obtained values for D and C from the Data Acquisition and Validation tasks, and showing how they affected the reliability value of the algorithm.

Substituting in the algorithm for operator reliability from the Data Acquisition task gives the following:

$$.9850 + [(.0150)(.6500)(.9231)] = .9940$$

The .6500 value for detection which was measured before the error light came on, can be partitioned into the following component values:

.3456 from cases when the subject signaled the detection to be intuitive,

.0265 from cases when the subject signaled the detection to be from feedback,

.2779 from cases when the subject failed to signal and went directly to the correction.

For cases when an intuitive detection was signaled, the obtained proportion of correction was .8596. For cases when a feedback detection was signaled, the obtained proportion of correction was .9444.

Substituting in the algorithm for operator reliability from the Validation task gives the following:

$$.9775 + [(.0225)(.6544)(.9503) = .9915$$

The .6544 value for detection, which is close to the detection value of .6500 for the Data Acquisition task, can be partitioned into the following values:

.2496 from cases when the subject signaled the detection to be intuitive,

.0251 from cases when the subject signaled the detection to be from feedback,

.3796 from cases when the subject failed to signal and went directly to correction.

For all cases when a intuitive detection was signaled, the obtained proportion of correction was .8824. Values for the Validation task closely followed the predictions from the Data Acquisition task.

The general conclusion from results of both tasks is that task reliabilty can be substantially reduced by operator detection and correction. For tasks similar to those used in the present study, the detection and correction values from this study could be used if no more appropriate data are available.

QUESTIONAIRE COMMENTS

Immediately following the last data collection trial, all subjects were asked to fill out a questionnaire giving their comments and criticisms regarding the present study. The questions asked of the subjects centered on several areas; namely, the training and data collection trials, use of the error keys, comments regarding the hardware equipment, and environmental conditions. While the comments given by all subjects were noted by the experimenter, no formal quantification of the comments was done.

There was total agreement by all subjects that the training and data collection trials were well designed and facilitated their learning of the task. Additionally, ample time was provided for breaks and lunch periods thus eliminating boredom or fatigue factors.

The next set of questions dealt with the use of the error keys, specifically asking the subjects if they were confused about which key to use or if they had ever used the wrong error key accidentally. Seventy percent of the subjects, on both tasks, commented that there was no confusion on their part with which error key to use. Twenty percent of the subjects said that initially there was some confusion deciding on which error key to use or that occasionally they used the wrong error key. All subjects in this category further stated that this problem was eliminated by becoming more familiar with the task. Two subjects (10%) however, stated that they experienced confusion discriminating between the two error keys throughout the entire task but on an intermittent basis. Following this question, subjects were asked if they experienced any additional problems associated with the error keys. Fifty percent of the subjects commented that on occasion they forgot to use either error key when an error was made. Similarly, 15 percent said that they had accidentally hit an error key when no error had been made.

The next set of questions dealt with the criteria the subjects had set for their performance. Forty percent of the subjects were trying to reduce their error rate as well as increase their speed. The remainder of the subjects were concentrating on one particular aspect, namely, reducing the error rate (35%) or decreasing their time scores (25%). Lastly, all subjects were asked if they felt that they could improve their performance scores; ninety percent of the subjects answered affirmatively.

In addition to the specific questions mentioned above, subjects were given the opportunity to comment on other aspects of the experiment. The majority of the subjects gave favorable comments regarding the touch entry CRT, such as extremely responsive and easier to use than a conventional keyboard. Several suggestions, such as adding a glare shield and being able to adjust the angle of the CRT were proffered.

SECTION 4

RESEARCH IMPLICATIONS

The three-dimensional model of HECAD Categories, Teichner's Classifications, and the Information Metric was shown to be successful in predicting performance reliability data, but not to be successful in predicting error detection and correction data. We are now warranted to evaluate the field validity of the performance reliability function of this model that has been demonstrated to be valid in the laboratory. This field evaluation could be accomplished by instrumenting an on-line operational console in a manner guided by the measures taken in the laboratory demonstration that has been reported.

In regard to error detection, it was encouraging that detection was positively correlated with performance reliability. Correlated variance is not random variance. Perhaps further investigation of the cueing events associated with a broader range of error events would lead to a basis for a prediction algorithm. However, it is not encouraging to have found such a low proportion of feedback detected errors compared to the high proportion of "intuitively" detected errors. This latter finding is suggestive of an idiosyncratic phenomenon rather than a nomothetic one.

On first thought it would seem as if the probability of correction for a task error should be predicted by the same variables that predict performance reliability. If not that, then one would suspect that correction probability should at least be positively correlated with performance reliability. Neither of these assumptions was confirmed in the present study. It would appear that correction probability may instead be related to the nature of events concomitant with and following the error process, such as the error response and the responses that follow it. This would suggest inventorying the types of errors that are made in specified situations, and attempting to base a prediction algorithm on the relationships established among such responses.

REFERENCES

Goldbeck, R. A., & Charlet J. D. Prediction of operator workstation performance (WDL-TR7071). Palo Alto: Aeronutronic Ford Corporation, WDL Division, November 1975.

Munger, S. J., Smith, R. W., & Payne, D. An index of electronic equipment operability: Data store. Pittsburgh: American Institute for Research, 1962.

Rabbitt, P. M. A. Error correction time without external error signals. Nature (Lond.), 1966, 212, 438. (a)

Rabbitt, P. M. A. Errors and error correction in choice response tasks. Journal of Experimental Psychology, 1966, 71, 264-272. (b)

Rabbitt, P. M. A. Three kinds of error-signaling responses in a serial choice task. Quarterly Journal of Experimental Psychology, 1968, 20, 179-188.

Rabbitt, P. M. A. Psychological refractory delay and response-stimulus interval in serial, choice response tasks. In Koster, W. (Ed.), Attention and Performance II. Amsterdam, the Netherlands: North Holland Publishing Company, 1969.

Rabbitt, P. M. A., & Vyas, S. M. An elementary preliminary taxonomy for some errors in laboratory choice RT tasks. Acta Psychologica, 1970, 33, 56-76.

Teichner, W. H. Quantitative models for predicting human visual/perceptual/motor performance. New Mexico State University: Human Performance Laboratory, 1974.

Teichner, W. H., & Williams, E. An information processing theory of Human performance. New Mexico State University: Human Performance Laboratory, 1979.